Beyond genome wide association studies in celiac disease by exploring the non-coding genome

de Almeida, Rodrigo

PART I
Chapter 4

A Celiac Disease-associated SNP in the LPP locus affects expression of the long non-coding RNA LPP-AS1

Manuscript in preparation

Rodrigo Almeida[1], Isis Ricao-Ponce[1], Patrick Deelen[1], Yang Li[1], Vinod Kumar[1], Yvonne Kooy-Winkelaar[2], Jeroen van Bergen[2], Frits Koning[2], Cisca Wijmenga[1] and Sebo Withoff[1].

[1]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands.
[2]University of Leiden, Leiden University Medical Center, Department of Immunohematology and Blood Transfusion, Leiden, the Netherlands.

Abstract

Celiac disease (CeD) is a common autoimmune disease affecting ~1% of the general population and in which genetically susceptible individuals develop a strong intestinal inflammatory reaction to dietary gluten. Of the 58 genetic factors associated with the disease to date, the most significant non-HLA association signal maps to intron 2 of the LPP gene. We applied a fine-mapping strategy to reduce the original 70 kb linkage disequilibrium block down to a region of 2.8 kb and pinpointed a functional regulatory variant in this region that intersects with several enhancers. Recent meta-analyses identified a SNP in the same haplotype that displays an expression quantitative trait locus (eQTL) effect on the neighboring BCL6 gene. To investigate whether a functional SNP in this region affects the expression of nearby transcripts, we performed eQTL mapping using 629 blood samples. We identified the CeD-associated SNP rs2030519 as exhibiting an eQTL effect on the long non-coding RNA (lncRNA) LPP-AS1. Using an RNA-seq dataset encompassing 1,474 libraries from 16 different tissues, we found that the expression of LPP, BCL6 and LPP-AS1 is intercorrelated, with the correlation strongest between LPP and BCL6 (correlation coefficient = 0.6). Subsequent pathway analysis uncovered the involvement of LPP and BCL6 in immune-related pathways and suggests that LPP-AS1 has a role in the ubiquitination pathway. Although we cannot exclude the possibility of BCL6 and LPP being involved in CeD, we have shown, for the first time, that a lncRNA (LPP-AS1) might play a role in celiac disease.

Introduction

Celiac disease (CeD) is characterized by a strong autoimmune inflammatory response to dietary gluten in individuals who carry the genetic susceptibility profile (1). Genome-wide association studies (GWAS) and fine-mapping approaches have identified 40 non-HLA loci to be associated with CeD (2–4). The strongest non-HLA association signal maps to the single nucleotide polymorphism (SNP) rs2030519, located in a 70 kb linkage disequilibrium (LD) block, and more specifically, in intron 2 of the LPP gene (chr3: 188069360–188139629, NCBI build 37). Recently, we performed a fine-mapping study on this locus, which led to the prioritization of rs4686484 as the functional SNP in this region (5). Although this data suggests that rs4686484 intersects with an enhancer sequence utilized in B and T-cells, this needs to be confirmed experimentally.

The LPP locus has also been associated with generalized vitiligo (6) and allergic responses (7, 8). The LPP gene is thought to function in cell proliferation, cell motility and cell-cell adhesion. The LPP protein has been described as playing a role in focal adhesion architecture and it has been suggested that it might

act as a transcription factor (9). LPP knockout mice show developmental and reproductive abnormalities, and embryonic fibroblasts derived from these mice display reduced migration capacity, reduced viability, and reduced expression of some Lpp interaction partners (10). It was suggested that reduced levels of LPP in the nucleus contribute to CeD, based on experiments performed in fibroblasts obtained from CeD patients versus those obtained from control subjects (11).

Recently, 2 different allergy meta-analyses using microarray data uncovered a role for LPP in allergy (7, 8). Expression quantitative trait locus (eQTL) mapping in B-cell and monocyte datasets found that SNPs rs9865818 and rs9864529 downregulate the expression of the BCL6 mRNA (7, 8). Both these SNPs are in strong LD with SNP rs2030519 ($r2 > 0.8$, $D' = 1$). The BCL6 gene, located ~650 kb upstream of the LPP gene, encodes a zinc finger transcriptional repressor that is also expressed in B cells and CD4+ T cells present in germinal centers (12). Mice with a BCL6 knockout develop an inflammatory response in multiple organs, characterized by infiltrations of eosinophils and IgE-bearing B lymphocytes typical of a Th2-mediated hyper-immune response (13).

To shed light on how SNPs in the LPP locus could contribute to CeD, we first performed eQTL mapping in a large RNA-sequencing (RNA-seq) dataset obtained from 629 in full (PBMC) samples isolated from subjects in the Lifelines Deep cohort (14). Remarkably, we found SNP rs2030519 to have a downregulation effect on the long non-coding RNA (lncRNA) LPP-AS1.

LncRNAs regulate gene expression in many biological processes, including the immune response (15). For instance, lincRNA-Cox2 was recently found to mediate both the activation and repression of distinct classes of immune genes (16). Moreover, lncRNAs are dysregulated in a variety of complex diseases and are closely associated with disease development and progression (17), for example in multiple cancers (18) and Alzheimer disease (19). One characteristic of lncRNAs is that they are expressed in a more tissue-specific manner than protein or coding genes (20).

In this study we set out to explore the expression patterns of the lcnRNA LPP-AS1, as well as of the two protein-coding genes (LPP and BCL6). First, we explored 1,474 publically available RNA-seq samples from specific cells or tissues likely to be related to CeD (Table 1). Further, we investigated RNA-seq data which we generated from a subset of seven immune cell types and gluten-specific T cell clones. We found LPP-AS1 had lower expression in immune cells and in the gluten-specific T cell clones compared to the protein-coding genes in both RNA-seq datasets (public data and our own). On the other hand, the

protein-coding genes were normally expressed across all the cells types. Next, we performed correlation analysis of these transcripts only in RNA-seq data from immune cells, which showed a strong correlation between LPP and BCL6 (correlation coefficient = 0.6), and with LPP-AS1 showing a weak correlation with the two protein-coding genes (correlation coefficient = 0.3).

We also performed enrichment pathway analysis based on genome-wide co-expression data from RNA-seq, which suggested immune and inflammatory pathways for the protein-coding genes, and a regulatory mechanism, such as ubiquitination, for the lncRNA. This is in agreement with the cell-type-specific expression data and can thus contribute to CeD pathogenesis. Our results suggested that the LPP gene itself might not be the major transcript in this CeD-associated locus and we cannot exclude BCL6 from also being implicated in the disease. Thus, our analysis suggested, for the first time, that a lncRNA (LPP-AS1) is probably involved in CeD etiology.

Methods

RNA-seq analysis

Raw sequencing reads were mapped to human reference genome NCBI, build 37, using STAR v2.1.3 (21), allowing for eight mismatches and five mapping positions. To reduce reference mapping bias, SNPs from the Genome of the Netherlands study (GoNL) (22) with MAF > 1% were masked by “N”. On average, 92% of reads were mapped and 88% of all reads were mapped uniquely. 88% of all aligned reads mapped to exons.

Gene expression was estimated using HTSeq count (23), using Ensembl GRCh37.71 gene annotation. Only uniquely mapping reads were used for estimating expression. Then we used gene expression data that was quantile normalized and log2-transformed for further analysis. The expression of each gene was centered and scaled. To reduce the effect of non-genetic sources of variability, we applied principal component analysis (PCA) on the sample correlation matrix and used the first ten components as covariates (24).

eQTL mapping

As a discovery set, we investigated 629 RNA-seq samples from the LifeLines Deep cohort (14) to map cis-eQTLs to 48 SNPs that had been associated to CeD by Immunochip analysis with genome-wide significance (3). Further, conditional analysis on the most associated SNP was performed to prioritize SNPs with an eQTL effect in the same LD block.

For trans-eQTLs, we performed a meta-analysis of the LifeLines Deep samples (14) and 338 samples from the GEUVADIS consortium dataset (25). Cis-eQTL analysis was performed on SNP-gene combinations for which the distance from the center of the gene to the genomic location of the SNP

was ≤ 250 kb, while eQTLs with a distance greater than 5 Mb were defined as trans-eQTLs. A weighted z-score method was used for further analysis. Associations were tested by the non-parametric Spearman's rank correlation test and we defined false discovery rate (FDR) significance thresholds based on the number of SNPs tested.

Cell-type-specific expression profiles

Publically available RNA-seq data for tissues and primary cell lines (1,474 samples) were downloaded from the European Nucleotide Archive (ENA) database (http://www.ebi.ac.uk/ena/). The quality of the RNA-seq data was checked as described previously (26). In short, PCA was performed to remove outliers and quantile normalization was applied across all samples (26). Unsupervised clustering, as well as the visualization of the expression data, was performed using the Pheatmap R-package v.07.7 (27). Co-expression analysis was performed by applying Spearman's rank correlation testing.

Collection of granulocytes and PBMC fractions

Two volunteers who signed informed consent forms (as required by the Institutional Review Board of the University Medical Centre Groningen) donated blood in lithium-heparin vacutainers (BD Biosciences, San Jose, CA, USA). The granulocyte/red blood cell fraction and PBMCs were separated using Ficoll Paque Plus (GE Healthcare Life Sciences, Diegem, Belgium), according to the manufacturer's instructions. The red blood cells in the granulocyte/red blood cell pellet were lysed in 155 mM NH4Cl, 10 mM KHCO3, 0.1 mM Na2, EDTA.2H2O, pH 7.4, to yield the granulocyte fraction. Monocytes were gated and sorted based on differences in forward and sideways scatter profiles. The lymphocyte fraction was further separated into B cells (CD4-, CD8-, CD19+), NK cells (CD4-, CD8-, CD58/CD16+), naive CD4 cells (CD4+, CD45RO-), naive CD8 cells (CD8+, CD45RO-) and memory T cells (CD4+, CD45RO+ or CD8+, CD45RO+) by flow sorting (MoFloTM XDP, Beckman Coulter, Wilsonville, Oregon, USA). Anti-CD8a-APC-eF780 and anti-CD4-eF450 were obtained from eBioscience (San Diego, CA, USA), anti- CD45RO-FITC and anti-CD19-AF700 from BD Biosciences, and anti-CD56-Pe and anti-CD16-Pe from IQP (Groningen, the Netherlands).

Gluten-specific T cell clones

Polyclonal gluten-specific T-cell lines were isolated from the small intestine of patients with CeD as described previously (28). In short, biopsies were cultured with a mixture of gluten and TG2-treated gluten for 5 days. Next, IL-2 (20 Cetus units/ml; Novartis, Arnhem, The Netherlands) and IL-15 (10 ng/ml; R&D systems, Abingdon, UK) were added to expand the T cells. Re-stimulation was performed with mixed, irradiated, allogeneic PBMCs

in the presence of phytohemagglutinin (1 g/ml; Remel Inc. Lenexa, USA), IL-2 Ja(20 Cetus units/ml) and IL-15 (10 ng/ml). The resulting T-cell lines were stored in liquid nitrogen and later tested for reactivity against a pepsin and trypsin digest of gluten, and against a pepsin and trypsin digest of gluten treated with TG2 in a T-cell proliferation assay. Gluten-reactive lines were cloned by limiting dilution and were again expanded by re-stimulation at 1 to 3-week intervals. Clones were stored in liquid nitrogen until thawed and stimulated with 2.5 ug/ml anti-CD3 and anti-CD28 (both obtained from Biolegend, ITK Diagnostics, Uithoorn, the Netherlands).

RNA isolation and library preparation

RNA was isolated from the purified immune cell populations using the Ambion mirVana kit (Life Technologies, Bleiswijk, the Netherlands), according to the manufacturer's instructions. The Nanodrop 1000 spectrometer (Thermo Fisher Scientific, Landsmeer, the Netherlands) and Experion High-sensitivity RNA analysis kit (Bio-Rad, Waltham, MA, USA) were used to determine RNA quantity and quality, respectively. Stranded RNA-seq libraries were prepared from 0.5-1 ug of ribosomal RNA-depleted total RNA, using the Ribo-Zero rRNA removal kit from Epicentre (Madison, Wisconsin, USA) and Illumina's Truseq Stranded Total RNA Sample Preparation Kit (San Diego, CA, USA), following the manufacturer's instructions, and subsequently sequenced on an Illumina HiSeq 2500 sequencer.

RNA-seq analysis of seven immune cells and gluten-specific T cell clones

Raw sequencing reads from the seven immune cells were mapped using STAR v2.1.3 (21) with the same settings as previous described here. For the immune cells (Supplementary Table 1), reads per kilobase per million (RPKM) were used for normalization. The following formula was used to calculated RPKM:

$$RPKMg = 109*(Cg/N*Lg)$$

where Cg is the number of reads that map into the exons of the gene g, Lg is the length of the gene's g exons, and N is the total number of mapped reads for each particular sample (29). For the normalization of the gluten-specific T cell data, the DESeq package (30) was applied. We performed RNA-seq expression in these cells at different time points (10, 30 and 180 minutes).

Pathway enrichment analysis

Pathway enrichment analysis was performed to discover possible biological pathways involved with the candidate transcripts. First, we applied the Genefriends v3.0 RNA-seq tool using the individual transcripts as the reference transcript (31). This tool uses a genome-wide co-expression map based on ~4,000 publicly available RNA-seq samples, obtained from the Sequence Read Archive (SRA) database (http://www.ncbi.nlm.

nih.gov/sra). Moreover, Genefriends integrates co-expression data with the widely used tool, DAVID (Database for Annotation, Visualization and Integrated Discovery v6.7;(32)) to improve functional enrichment analysis. We only considered pathways when three criteria were met: (1) an enrichment score ≥ 1.3; (2) an enrichment of ≥ 1.5-fold; and (3) a FDR of < 0.05.

Results

eQTL mapping in RNA-seq data

By performing eQTL mapping in 629 RNA-seq samples from the Lifelines Deep cohort, we found that the strongest non-HLA CeD-associated SNP, rs2030519 in the LPP gene, displayed an eQTL effect (P = 2.5 x 10-5, Z-score = -4.21) on lncRNA LPP-AS1 (Fig. 1).

Cell-type-specific expression profiles on public available RNA-seq data

To explore the expression of the candidate transcripts, we first investigated 1,474 publicly available RNA-seq samples of primary cells and tissues (Table 1). We found the BCL6 gene expressed in most cell types, including monocytes, B and T cells, and it was slightly more highly expressed in neutrophils (Fig. 2A). The LPP gene was also ubiquitously expressed, including B and T cells, and showed slightly higher expression in the large and small intestine (Fig. 2B). Both protein-coding genes showed lower expressed in lymphoblasts (Fig. 2A and 2B). Nevertheless, excepted for blood and monocytes, the lncRNA LPP-AS1 was found to have low expression in all other cells and tissues (Fig. 2C).

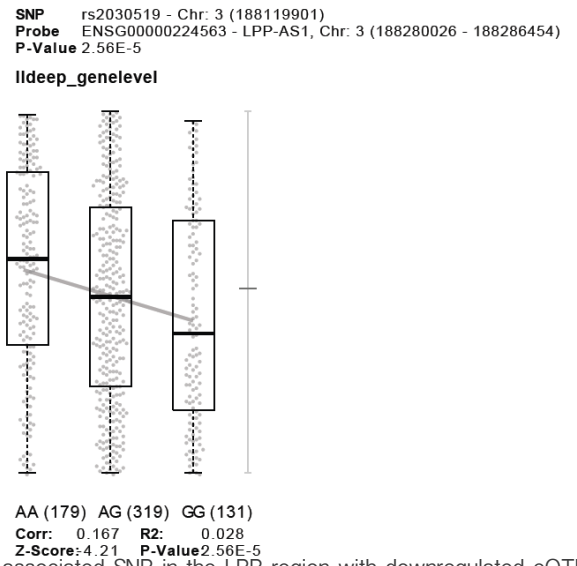


Figure 1. Most associated SNP in the LPP region with downregulated eQTL effect for lncRNA LPP-AS1.

Table 1. Description of cells and tissues from RNA-seq public data

Tissues	Number of samples
Lymphoblasts	1010
Blood	140
PBMCs	83
B lymphocytes	75
Hematopoietic cells	58
Epithelium	35
Large intestine	19
Small intestine	16
Bone marrow	14
T lymphocytes	7
Thymus	6
Monocytes	4
Neutrophil	4
Granulocytes	2
NK cells	1
Total	1474

In order to verify if these transcripts could be co-expressed with each other in this subset of cells and tissues, we performed a non-parametrical Spearman's correlation test between each transcript in the RNA-seq samples. We found that the lncRNA LPP-AS1 was weakly correlated with both candidates genes, showing correlation coefficients (cor) = 0.35 and P-value = 2.2 x 10-16 for BCL6; and cor = 0.34, P-value = 2 x 10-16 for LPP genes. Nevertheless, both the BCL6 and LPP genes were highly correlated (cor = 0.67 and P-value = 1.2 x 10-15) (Table 2).

To verify if these transcripts could also cluster differently according to their expression pattern, we performed unsupervised clustering in the same subset of RNA-seq samples. We found that both genes were in the same cluster, whereas LPP-AS1 was in a different cluster (Fig. 3), which agrees with our correlation analysis.

Expression profiling in seven immune cells and in gluten-specific T cells

To validate our findings using the publically available RNA-seq data, but also to obtain expression profiling in cell types that are known to be specifically involved in CeD, we performed RNA-seq on seven different peripheral blood-derived immune cell-types (Supplementary Table 1) and in gluten-specific T cells (Fig. 4). This analysis

Table 2. Correlation between the three transcripts

Transcripts	Correlation coefficient	P-value
LPP and BCL6	0.67	1.2×10^{-15}
LPP and LPP-AS1	0.34	2×10^{-16}
LPP-AS1 and BCL6	0.35	2.2×10^{-16}

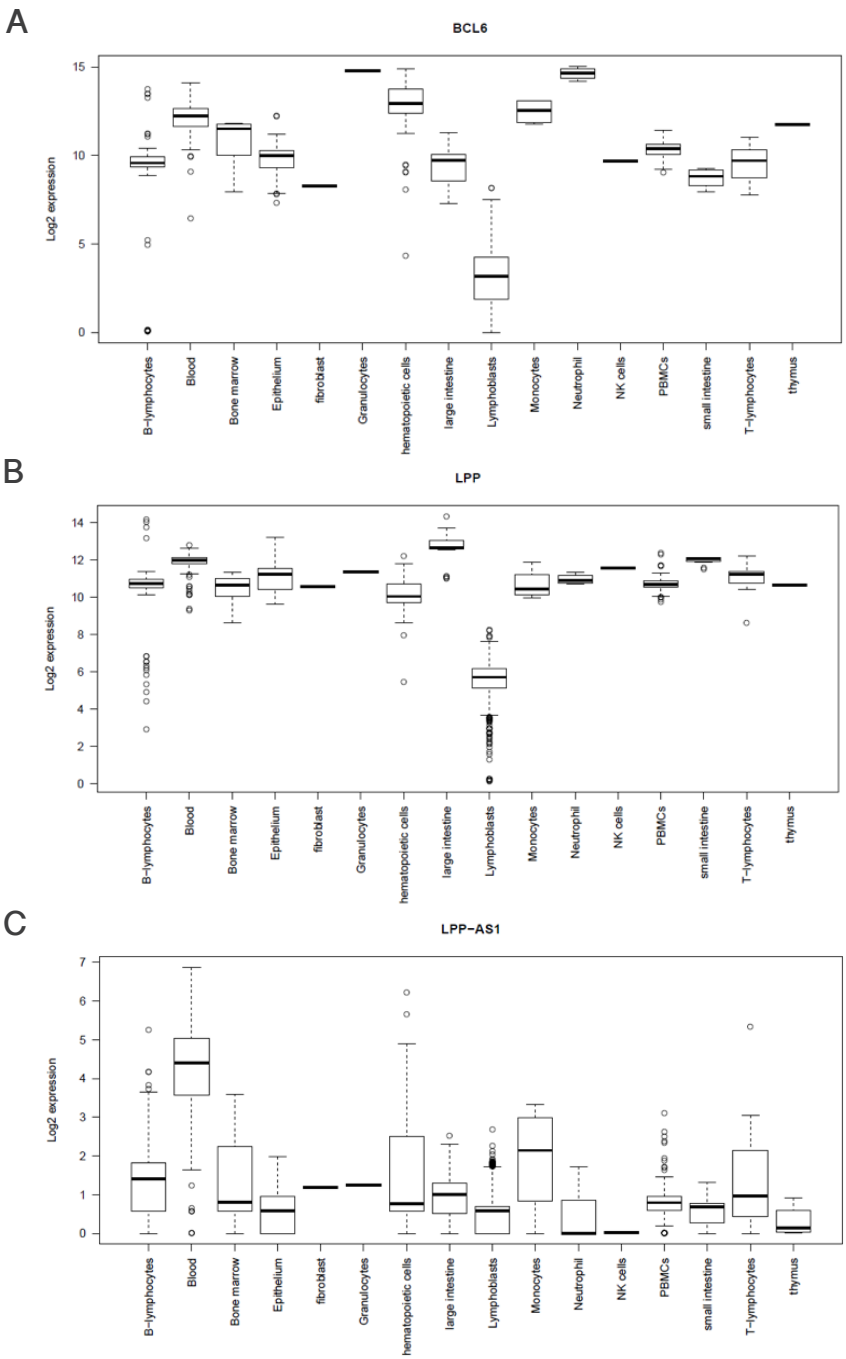


Figure 2. Distribution of expression data of 1,474 RNA-seq public data on the 3 transcripts. Gene expression data was quantile normalized and log2-transformed. (A) BCL6 is normally expressed across all other cells/tissues. (B) LPP was normally expressed in most of cells/tissues, where in lymphoblast was low expressed. (C) LPP-AS1 displayed low expression pattern in all cells/tissues excepted in blood.

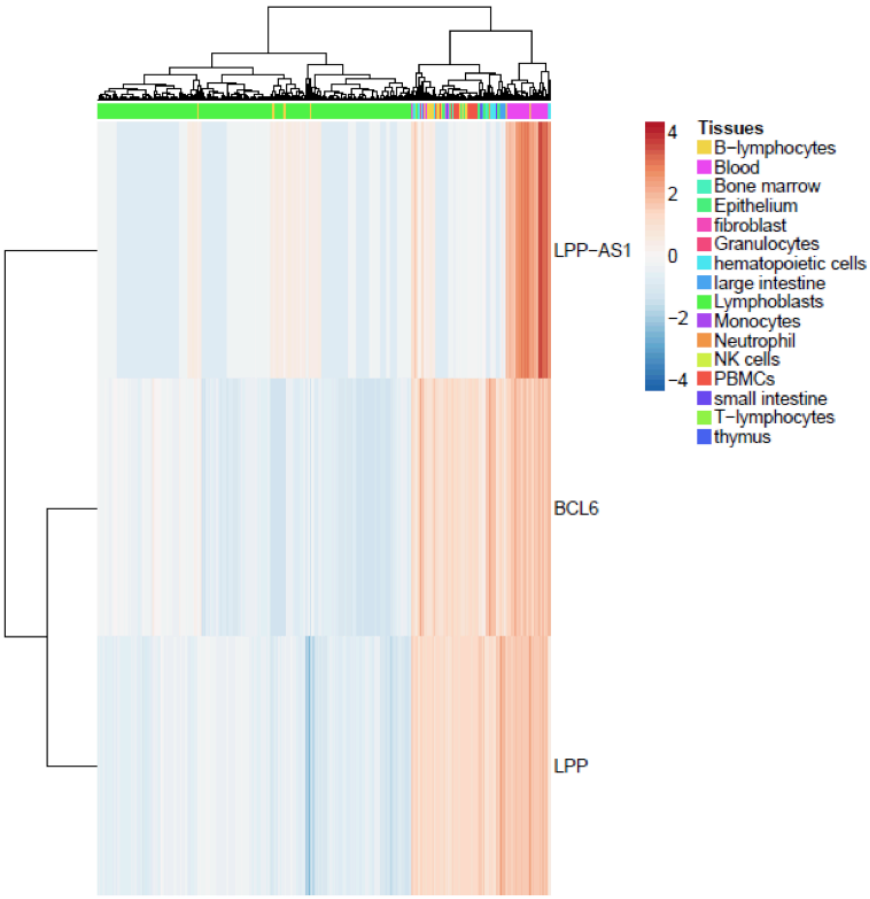


Figure 3. Unsupervised clustering of 1,474 specific cells and tissues. On top of the heatmap is the clustering of each of the 16 different tissues. LPP and BCL6 genes are more related to each other than the lncRNA LPP-AS1, which clusters in another branch. The two genes shown a similar pattern of expression across tissues. The lncRNA displayed high expression in blood.

confirmed the rather low expression of LPP-AS1 in immune cells and showed that BCL6 expression is highest in granulocytes, monocytes and B cells (Fig. 5). The LPP gene shows lower expression than BCL6 in the seven immune cells, but at higher levels than LPP-AS1 (Fig. 5). Additionally, by exploring stimulated gluten-specific T cells, we observed that BCL6 seems to be induced slightly, while LPP and LPP-AS1 levels remain stable (Fig. 4A and 4B), and the lncRNA showed low expression over the time points (Fig. 4C).

Pathway analysis in co-expression data from RNA-seq

In order to discover the biological pathways that these transcripts might be involved in, we performed pathway analysis based on co-expression data of ~2,500 RNA-seq samples (31). We found 161 pathways enriched in BCL6 (Supplementary Table 2), including several immune and inflammatory

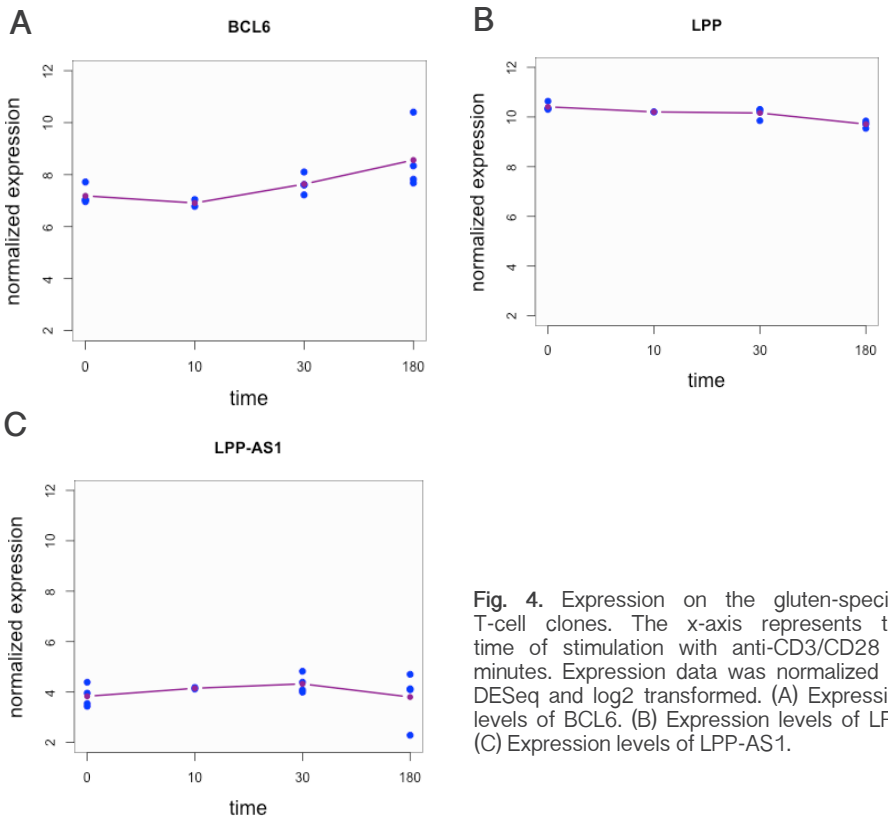


Fig. 4. Expression on the gluten-specific T-cell clones. The x-axis represents the time of stimulation with anti-CD3/CD28 in minutes. Expression data was normalized by DESeq and log2 transformed. (A) Expression levels of BCL6. (B) Expression levels of LPP. (C) Expression levels of LPP-AS1.

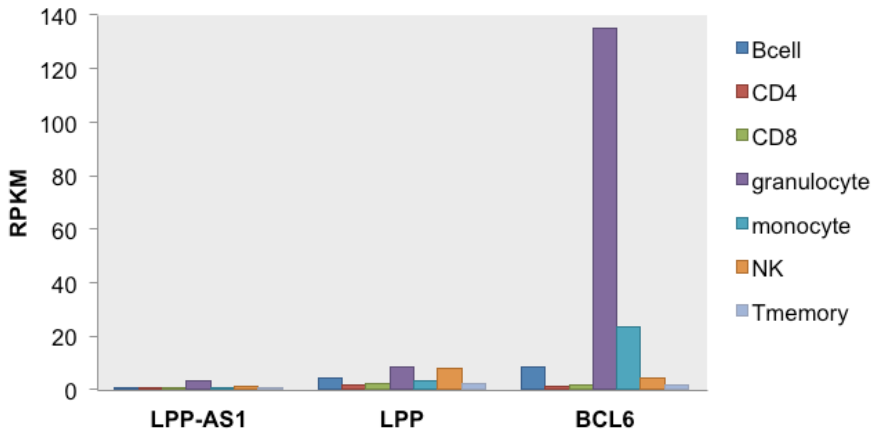
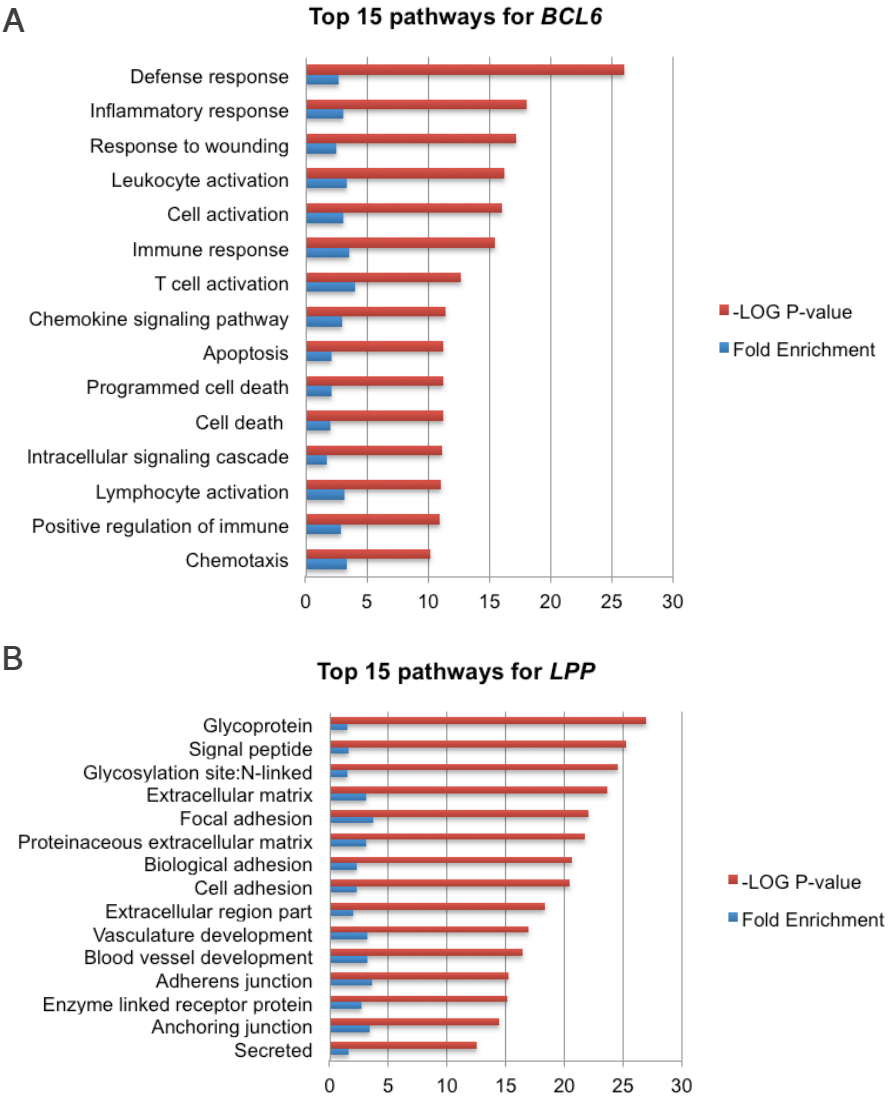


Figure 5. Expression profile of two volunteers. Expression levels in the y-axis normalized by reads per kilobase per million (RPKM). Low expression levels of LPP-AS1 in seven immune cell types and high expression levels of BCL6 in granulocytes.

pathways, such as leukocyte and T-cell activation (P = 6 x10-17 and P = 2.3 x10-13, respectively) (Fig. 6A). In addition, for the LPP gene we found 189 significant pathways (Supplementary Table 3), including focal adhesion (P = 8.6 x 10-23) and cell adhesion (P = 3.8 x 10-21) (Fig. 6B). For the lncRNA LPP-AS1, we found 18 significant pathways (Supplementary Table 4). Although most of these were related with reproduction in this lncRNA (Fig. 7C), we also found the ubiquitin-mediated proteolysis pathway (P = 0.002), which plays a role in several basic cellular processes, including modulation of immune and inflammatory responses (33).



C

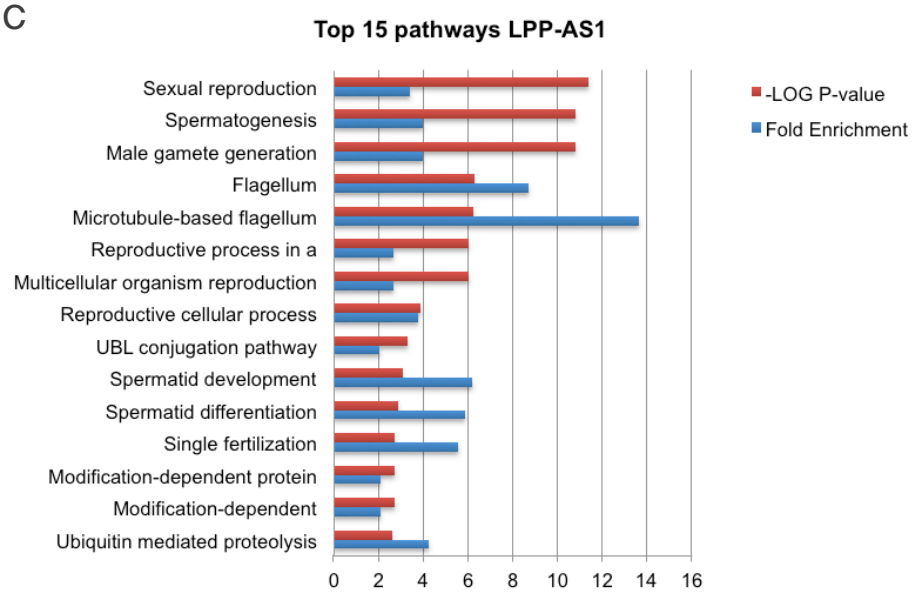


Figure 6. Pathway enrichment analysis. Fold enrichment and P-values were obtained by integrating the Genefriends and DAVID tools and are represented in each x-axis. Red bars are minus Log P-values and blue bars are the fold enrichment. (A) Top 15 pathways for BCL6 gene. (B) Top 15 pathways for LPP gene. (C) Top 15 pathways for LPP-AS1.

Discussion

By applying a fine-mapping approach we were recently able to reduce the size of the CeD-associated LD block of the LPP locus from 70 to 2.8 kb. Our analysis also indicated B and T cell-specific enhancer regions that might be affected by functional SNPs in this smaller region (5). Almost simultaneously, two meta-analysis studies showed that two SNPs present in our finely mapped LPP LD block exhibited an eQTL effect that downregulates BCL6 in B-cells and monocytes, respectively (7, 8). In order to obtain more information on the potential involvement of additional genes in this locus, we decided to analyze the RNA-seq datasets we had available. By performing eQTL mapping using PBMC samples from 629 individuals, we uncovered an eQTL effect of rs2030519 (the most associated SNP in the LPP region), downregulating lncRNA LPP-AS1 and, for the first time, implicating a lncRNA as involved in CeD.

By exploring additional RNA-seq datasets of primary cells and tissues, we found LPP-AS1 to be less but broadly expressed in cellular and tissue subsets, whereas the two protein-coding genes were expressed at a higher level. It should be noted that, in general, fluctuation in lncRNA levels is more difficult to determine, because of their lower expression.

Initially, the LPP gene was considered to be the primary causal gene associated with this locus, and interestingly we observed that the LPP

gene was expressed slightly higher in intestinal cells. This is in agreement with our recent findings, which suggest a role for LPP in the intestinal barrier function (34). Moreover, we observed that the expression level of LPP was decreased in intestinal biopsies from severely affected CeD patients compared with healthy individuals (5). Preliminary in vitro barrier function assay data supports a role for LPP in the barrier function (Maria Zorro and Sven van Ijzendoorn, personal communication). Our pathway enrichment analysis confirmed co-expression of LPP with genes involved in focal adhesion and cell adhesion pathways. These pathways are critical for maintaining the intestinal epithelial barrier function (35). We did not find an eQTL effect on the LPP gene itself in the PBMC dataset, which could mean that either LPP is not the causal gene in the locus, or that we haven't investigated the specific cell type in which the function of LPP is most critical. Importantly, we cannot rule out the possibility that there is an epithelial-cell-specific eQTL effect on LPP. Unfortunately, we could not interrogate this possibility in the data available presently.

Although the BCL6 gene is located 658.7 kb away from the CeD LPP LD region, genetic variants were found in the LPP LD block that exhibit an eQTL effect (7, 8) on BCL6, suggesting that BCL6 might play a role in CeD. Recently, a similar observation was described for a SNP in the FTO gene associated with type 2 diabetes (36) and obesity (37). This particular intronic SNP was reported to have an eQTL effect on the IRX3 gene located megabases away from the FTO locus (38). We tried to reproduce the eQTL effect on BCL6 in our RNA-seq dataset, but did not succeed. A reason for this could be that the trans-eQTL effect observed on this gene is cell-type-specific. In fact, it has been reported that only 7% of trans-eQTLs are shared between B cells and monocytes and it is quite likely that the relevant cell type is under-represented or even absent in the PBMC samples we used (39). Additionally, the sample size could have been too small to identify more subtle trans-eQTL effects of these SNPs. Our pathway enrichment analysis confirmed that BCL6 is involved in immune-related pathways.

LncRNA eQTLs have not been explored extensively thus far. The published meta-analysis eQTL studies used microarray data (7, 8), so lncRNAs – including LPP-AS1 – were not queried. Our study identified the first lncRNA involved in CeD and very recently, we have described GWAS SNPs with eQTL effects on lncRNAs (40). Some of the disease-associated lncRNAs identified in this study are co-expressed with neighboring protein-coding genes, suggesting a function for the lncRNA in regulating the expression of the protein-coding gene. In our data we did not find a correlation between LPP-AS1 expression and the expression of BCL6 or LPP. To formally prove that LPP-AS1 is not

regulating BCL6 and/or LPP, LPP-AS1 should be knocked down or deleted and the transcriptional consequences of this should be investigated.

Pathway enrichment analysis for LPP-AS1 suggests a role for this lncRNA in the ubiquitination pathway, which was previously implicated in CeD, also by application of a pathway prediction approach (41). Previous GWAS and Immunochip analysis in our lab has associated ubiquitination genes with CeD. It is known that lncRNAs may serve as modular scaffolds for assembling multi-protein complexes (42). Interestingly, HOTAIR has been described as forming complexes with two mouse E3 ubiquitin ligases (Dzip3 and Mex3b) and with their respective substrates (Ataxin-1 and Shurportin-1). This suggests that HOTAIR functions as a scaffold transcript that facilitates the E3-mediated ubiquitination of substrate proteins (43). It is possible that that LPP-AS1 plays a similar role in ubiquitination and that the involvement of lncRNAs in ubiquitination is a common theme. Dysregulated LPP-AS1 function could be linked to dysregulating the immune response in CeD and autoimmune disease, in general, as ubiquitination is involved in multiple immune signaling pathways, such as NF-kB (44) and TGF-beta signaling (45). It will need functional studies in the specific cell types in which the lncRNAs are expressed (e.g. B and T cells) to prove these hypotheses.

In summary, although our findings do not rule out the possibility that the LPP or BCL6 genes may be involved in the disease, we did find evidence that implicates LPP-AS1 as the first lncRNA involved in CeD.

Acknowledgements

We thank Jackie Senior for editing this chapter and Javier Gutierrez-Achury for discussion and help with statistical analyses. We thank the Target project (http://www.rug.nl/target) for providing the compute infrastructure. This work was supported by funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement (no. 2012-322698) to C.W., the Dutch Digestive Disease Research Foundation (MLDS, WO11-30 to C.W.), and the Netherlands Organization for Scientific Research (NWO-VICI grant 918.66.620) to C.W.

References

1. Guandalini, S. and Assiri, A. (2014) Celiac disease: a review. JAMA Pediatr., 168, 272–8.

2. Dubois, P.C. a, Trynka, G., Franke, L., Hunt, K. a, Romanos, J., Curtotti, A., Zhernakova, A., Heap, G. a R., Adny, R., Aromaa, A., et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet., 42, 295–302.

3. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-rosun, L., Castillejo, G., et al. (2011) Articles Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet., 43, 1193–1201.

4. Hunt, K. a, Mistry, V., Bockett, N. a, Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O., et al. (2013) Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature, 498, 232–5.

5. Almeida, R., Ricao-Ponce, I., Kumar, V., Deelen, P., Szperl, A., Trynka, G., Gutierrez-Achury, J., Kanterakis, A., Westra, H.-J., Franke, L., et al. (2014) Fine mapping of the celiac disease-associated LPP locus reveals a potential functional variant. Hum. Mol. Genet., 23: 2481-9.

6. Jin, Y., Birlea, S.A., Fain, P.R., Ferrara, T.M., Ben, S., Riccardi, L., Cole, J.B., Gowan, K, Holland, P.J., Bennett, D.C., et al. (2012) NIH Public Access. 44, 676–680.

7. Bnnelykke, K, Matheson, M.C., Pers, T.H., Granell, R., Strachan, D.P., Alves, A.C., Linneberg, A., Curtin, J., Warrington, N.M., Standl, M., et al. (2013) Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat. Genet., 45, 902–6.

8. Hinds, D. a, McMahon, G., Kiefer, A.K., Do, C.B., Eriksson, N., Evans, D.M., St Pourcain, B., Ring, S.M., Mountain, J.L., Francke, U., et al. (2013) A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat. Genet., 45, 907–11.

9. Grunewald, T.G.P., Pasedag, S.M. and Butt, E. (2009) Cell Adhesion and Transcriptional Activity — Defining the Role of the Novel Protooncogene LPP. Transl. Oncol., 2, 107–116.

10. Vervenne, H.B.V.K, Crombez, KR.M.O., Delvaux, E.L., Janssens, V., Van de Ven, W.J.M. and Petit, M.M.R. (2009) Targeted disruption of the mouse Lipoma Preferred Partner gene. Biochem. Biophys. Res. Commun., 379, 368–73.

11. Nanayakkara, M., Kosova, R., Lania, G., Sarno, M., Gaito, A., Galatola, M., Greco, L., Cuomo, M., Troncone, R., Auricchio, S., et al. (2013) A celiac cellular phenotype, with altered LPP sub-cellular distribution, is inducible in controls by the toxic gliadin peptide P31-43. PLoS One, 8, e79763.

12. Ye BH, Lista F, Lo Coco F, Knowles DM, Offit K, Chaganti RS, D.-F.R. (1993) Alterations of a Zinc Finger-Encoding Gene, BCL-6, in Diffuse Large-Cell Lymphoma. Science., 262, 747–50.

13. Ye, B.H., Cattoretti, G., Shen, Q., Zhang, J., Hawe, N., de Waard, R., Leung, C., Nouri-Shirazi, M., Orazi, A., Chaganti, R.S., et al. (1997) The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. Nat. Genet., 16, 161–170.

14. Ettje F. Tigchelaar, Alexandra Zhernakova, Jackie A.M. Dekens, Gerben Hermes, Agnieszka Baranska, Zlatan Mujagic, Morris A. Swertz, Anglica M. Muoz, Patrick Deelen, Maria C. Cnit, Lude Franke, Salome Scholtens, Ronald P. Stolk,

Cisca Wijmenga, E.J.M.F. (2014) An introduction to LifeLines DEEP: study design and baseline characteristics. (Manuscript in preparation).

15. Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. Annu. Rev. Biochem., 81, 145–66.

16. Carpenter, S., Atianand, M., Aiello, D., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., et al. (2013) A Long Noncoding RNA Mediates Both Activation and Repression of Immune Response Genes. Science, 341:789-92.

17. Hrdlickova, B., de Almeida, R.C., Borek, Z. and Withoff, S. (2014) Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochim. Biophys. Acta, 1842, 1910-22.

18. Gibb, E.A., Vucic, E.A., Enfield, K.S.S., Stewart, G.L., Lonergan, K.M., Kennett, J.Y., Becker-Santos, D.D., MacAulay, C.E., Lam, S., Brown, C.J., et al. (2011) Human cancer long non-coding RNA transcriptomes. PLoS One, 6, e25915.

19. Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St Laurent, G., Kenny, P.J. and Wahlestedt, C. (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat. Med., 14, 723–30.

20. Washietl, S., Kellis, M. and Garber, M. (2014) Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res., 24, 616–28.

21. Dobin, A., Davis, C. a, Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29, 15–21.

22. The Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet., 46, 818–825.

23. Anders, S., Pyl, P.T. and Huber, W. (2014) HTSeq A Python framework to work with high- throughput sequencing data Cold Spring Harbor Labs Journals.pii: btu638, [Epub ahead of print].

24. Fehrmann, R.S.N., Jansen, R.C., Veldink, J.H., Westra, H.-J., Arends, D., Bonder, M.J., Fu, J., Deelen, P., Groen, H.J.M., Smolonska, A., et al. (2011) Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet., 7, e1002197.

25. Lappalainen, T., Sammeth, M., FriedInder, M.R., 't Hoen, P. a. C., Monlong, J., Rivas, M. a., Gonzlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013) Transcriptome and genome sequencing uncovers functional variation in humans. Nature, 501:506-11.

26. Deelen, P., Zhernakova, D. and Haan, M. De (2014) Calling genotypes from public RNA-sequencing data enables identification of genetic variants that affect gene-expression levels.Cold Spring Harbor Journal. 007633.

27. Kolde, R. (2013) pheatmap: Pretty Heatmaps.(www.r-project.org).

28. Petersen, J., Montserrat, V., Mujico, J.R., Loh, K.L., Beringer, D.X., van Lummel, M., Thompson, A., Mearin, M.L., Schweizer, J., Kooy-Winkelaar, Y., et al. (2014) T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. Nat. Struct. Mol. Biol., 21, 480–8.

29. Mortazavi, A., Williams, B.A., McCue, K, Schaeffer, L. and Wold, B. (2008)

Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods, 5, 621–8.

30. Anders, S., Mccarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K, Huber, W. and Robinson, M.D. (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. 8: 1765-86.

31. Van Dam, S., Cordeiro, R., Craig, T., van Dam, J., Wood, S.H. and de Magalhes, J.P. (2012) GeneFriends: an online co-expression analysis tool to identify novel gene targets for aging and complex diseases. BMC Genomics, 13, 535.

32. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc., 4, 44–57.

33. Ciechanover, A., Orian, A. and Schwartz, A.L. (2000) Ubiquitin-mediated proteolysis: biological regulation via destruction. Bioessays, 22, 442–51.

34. Kumar, V., Gutierrez-Achury, J., Kanduri, K, Almeida, R., Hrdlickova, B., Zhernakova, D. V, Westra, H.-J., Karjalainen, J., Ricao-Ponce, I., Li, Y., et al. (2014) Systematic Annotation of Celiac Disease Loci Refines Pathological Pathways and Suggests a Genetic Explanation for Increased Interferon Gamma levels. Hum. Mol. Genet., pii: ddu453.

35. Clemente, M.G., De Virgiliis, S., Kang, J.S., Macatagney, R., Musu, M.P., Di Pierro, M.R., Drago, S., Congia, M. and Fasano, A. (2003) Early effects of gliadin on enterocyte intracellular signalling involved in intestinal barrier function. Gut, 52, 218–23.

36. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science, 316, 889–94.

37. Dina, C., Meyre, D., Gallina, S., Durand, E., Krner, A., Jacobson, P., Carlsson, L.M.S., Kiess, W., Vatin, V., Lecoeur, C., et al. (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. Nat. Genet., 39, 724–6.

38. Smemo, S., Tena, J.J., Kim, K-H., Gamazon, E.R., Sakabe, N.J., Gmez-Marn, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature, 507:371-5.

39. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K, Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O. and Knight, J.C. (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet., 44, 502–10.

40. Kumar, V., Westra, H.-J., Karjalainen, J., Zhernakova, D. V, Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., Vsa, U., et al. (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet., 9, e1003201.

41. Castellanos-Rubio, A., Santin, I., Martin-Pagola, A., Irastorza, I., Castao, L., Vitoria, J.C. and Bilbao, J.R. (2010) Long-term and acute effects of gliadin on small intestine of patients on potentially pathogenic networks in celiac disease. Autoimmunity, 43, 131–9.

42. Guttman, M. and Rinn, J.L. (2012) Modular regulatory principles of large non-coding RNAs. Nature, 482, 339–46.

43. Yoon, J.-H., Abdelmohsen, K, Kim, J., Yang, X., Martindale, J.L., Tominaga-Yamanaka, K, White, E.J., Orjalo, A. V, Rinn, J.L., Kreft, S.G., et al.

(2013) Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. Nat. Commun., 4, 2939.

44. Wertz, I.E. and Dixit, V.M. (2010) Signaling to NF-kappaB: regulation by ubiquitination. Cold Spring Harb. Perspect. Biol., 2, a003350.

45. Wohlfert, E.A., Gorelik, L., Mittler, R., Flavell, R.A. and Clark, R.B. (2006) Cutting edge: deficiency in the E3 ubiquitin ligase Cbl-b results in a multifunctional defect in T cell TGF-beta sensitivity in vitro and in vivo. J. Immunol., 176, 1316–20.

Supplemental information

Supplementary Table 1. Seven immune cells types used for RNA-seq

1.	NK-cells
2.	B-cell
3.	Monocytes
4.	Memory T-cells
5.	CD4 T-cells
6.	CD8 T-cells
7.	Granulocytes

Supplementary Table 2. Enrichment Pathways for BCL6 gene

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Immune response	4.38	1.25E-45	3.02	4.60E-42
Defense response	2.91	4.83E-30	2.69	1.78E-26
Inflammatory response	2.91	7.65E-22	3.05	2.83E-18
Response to wounding	2.91	7.12E-21	2.49	2.63E-17
Leukocyte activation	3.1	8.22E-20	3.3	3.04E-16
Cell activation	3.1	1.58E-19	3.06	5.84E-16
T cell activation	3.1	4.03E-16	4.02	1.64E-12
Apoptosis	2.91	8.87E-15	2.58	5.36E-12
Chemokine signaling pathway	2.02	2.56E-14	2.91	4.31E-12
Cell death	2.91	1.25E-14	2.01	4.60E-11
Programmed cell death	2.91	1.63E-14	2.1	6.03E-11
Intracellular signaling cascade	2.86	2.11E-14	1.71	7.80E-11
Lymphocyte activation	3.1	3.39E-14	3.11	1.25E-10
Positive regulation of immune system process	3.35	4.69E-14	2.89	1.74E-10
Chemotaxis	2.7	2.81E-13	3.31	1.04E-09
Taxis	2.7	2.81E-13	3.31	1.04E-09
Regulation of cytokine production	3.35	5.80E-13	3.11	2.15E-09
Regulation of apoptosis	2.91	1.03E-12	1.86	3.82E-09
Regulation of programmed cell death	2.91	1.01E-12	1.86	3.72E-09
Regulation of cell death	2.91	1.32E-12	1.85	4.90E-09
Innate immune response	3.35	1.70E-12	3.43	6.30E-09
Immunoglobulin domain	4.38	3.65E-11	2.17	2.20E-08
Cytokine receptor	3.1	3.20E-11	9.68	1.93E-08
Myeloid leukocyte activation	3.45	5.44E-11	5.39	2.01E-07
Response to molecule of bacterial origin	2.91	1.05E-10	3.93	3.87E-07
Protein kinase cascade	2.86	1.10E-10	2.22	4.07E-07
Regulation of cell activation	3.1	1.27E-10	2.9	4.71E-07
Positive regulation of programmed cell death	2.91	1.43E-10	2.11	5.29E-07
Positive regulation of cell death	2.91	1.80E-10	2.1	6.67E-07
Positive regulation of apoptosis	2.91	2.48E-10	2.1	9.15E-07
Positive regulation of cell activation	3.1	2.47E-10	3.45	9.13E-07

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Positive regulation of cytokine production	3.1	3.62E-10	3.75	1.34E-06
Transmembrane protein	4.38	8.28E-10	1.9	4.99E-07
Natural killer cell mediated cytotoxicity	2.02	1.39E-09	2.82	2.33E-07
Regulation of leukocyte activation	3.1	9.73E-10	2.85	3.60E-06
Response to bacterium	2.91	1.11E-09	2.68	4.09E-06
Nod-like receptor signaling pathway	2.91	5.35E-09	3.78	8.98E-07
Regulation of alpha-beta T cell activation	3.1	1.97E-09	5.63	7.27E-06
Fc gamma r-mediated phagocytosis	2.02	4.20E-09	3.16	7.06E-07
Positive regulation of immune response	3.35	2.43E-09	2.95	8.99E-06
Positive regulation of lymphocyte activation	3.1	2.60E-09	3.48	9.62E-06
Immunoglobulin subtype	4.38	5.28E-10	2.3	8.95E-07
Cellular defense response	1.42	4.48E-09	4.25	1.66E-05
Regulation of lymphocyte activation	3.1	4.46E-09	2.89	1.65E-05
Positive regulation of response to stimulus	3.35	4.40E-09	2.43	1.63E-05
Response to lipopolysaccharide	2.91	4.95E-09	3.8	1.83E-05
Positive regulation of leukocyte activation	3.1	5.84E-09	3.29	2.16E-05
Positive regulation of alpha-beta T cell activation	3.1	6.05E-09	6.5	2.24E-05
Death-like	2.91	3.80E-10	4.42	6.44E-07
Regulation of i-kappab kinase/NF-kappab cascade	2.91	7.44E-09	3.26	2.75E-05
Toll-like receptor signaling pathway	2.91	2.18E-08	2.97	3.67E-06
Regulation of protein kinase cascade	2.91	1.03E-08	2.35	3.80E-05
Positive regulation of lymphocyte differentiation	3.1	1.07E-08	5.47	3.94E-05
Immune effector process	3.45	1.25E-08	2.94	4.64E-05
Cytosol	2.86	4.36E-09	1.53	2.12E-06
Cytokine-mediated signaling pathway	3.1	1.57E-08	3.86	5.81E-05
Cell activation during immune response	3.45	1.76E-08	5.32	6.51E-05
Leukocyte activation during immune response	3.45	1.76E-08	5.32	6.51E-05
T cell differentiation	3.1	1.74E-08	3.99	6.43E-05
Regulation of T cell activation	3.1	1.83E-08	3.08	6.77E-05
Chemotaxis	2.7	2.44E-08	3.98	1.47E-05
Positive regulation of T cell differentiation	3.1	2.02E-08	5.63	7.47E-05
Sh2 domain	4.5	3.51E-08	3.25	2.12E-05
Regulation of alpha-beta T cell differentiation	3.1	3.73E-08	6.31	1.38E-04
Leukocyte migration	2.7	3.86E-08	4.15	1.43E-04
Domain: ig-like v-type	3.61	2.39E-09	3.29	8.37E-06
Positive regulation of t cell activation	3.1	8.90E-08	3.56	3.29E-04
Response to organic substance	1.47	1.02E-07	1.67	3.76E-04
Regulation of tumor necrosis factor production	3.1	1.09E-07	5.45	4.01E-04
Regulation of adaptive immune response	3.1	1.56E-07	4.02	5.77E-04
Regulation of T cell differentiation	3.1	1.67E-07	4.2	6.16E-04

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Kinase	2.86	2.17E-07	1.72	1.31E-04
Positive regulation of defense response	3.1	1.84E-07	3.55	6.79E-04
Response to cytokine stimulus	2.91	1.96E-07	3.42	7.23E-04
Positive regulation of alpha-beta T cell differentiation	3.1	2.03E-07	6.76	7.49E-04
Leukocyte chemotaxis	2.7	2.18E-07	4.87	8.07E-04
Sh2 motif	4.5	2.47E-08	3.2	4.19E-05
Immune response-activating signal transduction	3.1	2.34E-07	4.12	8.65E-04
Regulation of lymphocyte differentiation	3.1	2.57E-07	3.75	9.49E-04
Immunoglobulin-like fold	4.38	3.59E-08	1.84	6.08E-05
Phosphorylation	2.86	3.09E-07	1.61	1.14E-03
Protein amino acid phosphorylation	2.86	3.20E-07	1.67	1.18E-03
Cytokine-cytokine receptor interaction	4.38	7.80E-07	1.97	1.31E-04
Sh3 domain	3.06	6.43E-07	2.37	3.88E-04
Cell chemotaxis	2.7	4.94E-07	4.62	1.83E-03
Lytic vacuole	3.18	2.05E-07	2.36	9.95E-05
Lysosome	3.18	2.05E-07	2.36	9.95E-05
Hematopoietic cell lineage	3.51	1.61E-06	2.84	2.71E-04
Positive regulation of i-kappab kinase/nf-kappab cascade	2.91	7.61E-07	3.02	2.81E-03
Immune response-regulating signal transduction	3.1	8.26E-07	3.82	3.05E-03
Membrane organization	3.18	8.70E-07	1.89	3.21E-03
Domain: ph	3.1	5.55E-08	2.4	1.94E-04
Leukocyte differentiation	3.1	9.77E-07	2.67	3.61E-03
Cytokine production	2.91	1.02E-06	4.16	3.75E-03
Leukocyte mediated immunity	3.62	1.04E-06	3.14	3.83E-03
Positive regulation of tumor necrosis factor production	2.91	1.01E-06	7.51	3.73E-03
Binding site: atp	2.86	4.52E-08	1.87	1.58E-04
Igg-binding protein	3.51	1.73E-06	12.45	1.04E-03
Protein kinase, core	2.86	1.73E-07	1.87	2.93E-04
Positive regulation of protein kinase cascade	2.91	1.19E-06	2.43	4.38E-03
Serine/threonine-protein kinase	2.86	1.67E-06	1.93	1.01E-03
Actin cytoskeleton organization	1.75	1.21E-06	2.19	4.48E-03
Induction of apoptosis	2.91	1.27E-06	1.97	4.68E-03
Induction of programmed cell death	2.91	1.41E-06	1.97	5.19E-03
Activation of immune response	3.35	1.51E-06	3	5.58E-03
Locomotory behavior	2.7	1.50E-06	2.06	5.51E-03
Phosphate metabolic process	2.86	1.48E-06	1.5	5.46E-03
Phosphorus metabolic process	2.86	1.48E-06	1.5	5.46E-03
Vacuole	3.18	7.57E-07	2.16	3.68E-04
T cell receptor signaling pathway	2.02	4.72E-06	2.52	7.93E-04
Adaptive immune response	3.35	2.15E-06	3.22	7.91E-03
Response to virus	3.03	2.27E-06	2.79	8.34E-03

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Lymphocyte differentiation	3.1	2.52E-06	2.84	9.26E-03
Regulation of protein amino acid phosphorylation	2.91	2.79E-06	2.34	1.03E-02
Protein kinase, atp binding site	2.86	5.58E-07	1.85	9.46E-04
Pleckstrin homology	3.1	4.97E-07	2.14	8.42E-04
Regulation of leukocyte mediated immunity	3.1	3.33E-06	3.51	1.22E-02
Membrane invagination	3.18	3.68E-06	2.15	1.35E-02
Endocytosis	3.18	3.68E-06	2.15	1.35E-02
Toll-interleukin receptor	3.51	7.65E-07	6.19	1.30E-03
Pattern recognition receptor signaling pathway	2.91	4.16E-06	6.63	1.53E-02
Lysosome	3.18	6.58E-06	2.51	3.96E-03
Src homology-3 domain	3.06	1.16E-06	2.28	1.96E-03
Positive regulation of signal transduction	2.91	5.75E-06	1.95	2.10E-02
Regulation of immune effector process	3.1	5.91E-06	2.79	2.16E-02
Positive regulation of cell communication	2.91	6.55E-06	1.88	2.39E-02
Actin filament-based process	1.75	6.87E-06	2.06	2.51E-02
Lymphocyte activation during immune response	3.1	7.67E-06	6.26	2.79E-02
B cell receptor signaling pathway	4.5	2.10E-05	2.75	3.51E-03
Immunoglobulin-like	4.38	2.49E-06	1.75	4.21E-03
Aig1	1.49	2.31E-06	11.86	3.90E-03
Interleukin-1 receptor, type i/toll precursor	3.51	2.19E-06	8.21	3.70E-03
Jak-stat signaling pathway	3.1	2.44E-05	2.12	4.10E-03
Positive regulation of multicellular organismal process	3.1	9.48E-06	2.03	3.45E-02
Phagocytosis	2.28	1.00E-05	3.75	3.63E-02
Leukocyte transendothelial migration	2.02	2.87E-05	2.31	4.81E-03
Regulation of innate immune response	3.35	1.09E-05	3.55	3.96E-02
Surface antigen	2.97	1.53E-05	4.15	9.16E-03
Activation of innate immune response	2.91	1.34E-05	5.93	4.84E-02
Innate immune response-activating signal transduction	2.91	1.34E-05	5.93	4.84E-02
Regulation of cd4-positive, alpha beta t cell differentiation	3.1	1.37E-05	6.76	4.93E-02
Macrophage activation	2.91	1.37E-05	6.76	4.93E-02
Domain:tir	3.51	8.58E-07	6.19	2.99E-03
Serine/threonine protein kinase-related	2.86	4.86E-06	1.88	8.20E-03
Gtpase activation	3.1	2.40E-05	2.35	1.44E-02
Domain:sh3	3.06	1.72E-06	2.45	6.00E-03
Pleckstrin homology-type	3.1	6.91E-06	1.96	1.16E-02
Serine/threonine protein kinase, active site	2.86	6.64E-06	1.88	1.12E-02
Hemopoiesis	3.1	2.16E-05	2	7.66E-02
Domain:protein kinase	2.86	1.63E-06	1.82	5.69E-03
Regulation of inflammatory response	2.91	2.53E-05	2.96	8.92E-02
T cell receptor complex	3.1	1.86E-05	7.73	8.98E-03

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Immunoglobulin v-set	4.38	1.09E-05	2.05	1.83E-02
Caspase recruitment	2.91	1.22E-05	4.91	2.04E-02
Domain:sh2	2.14	6.18E-06	2.96	2.14E-02
Active site:proton acceptor	2.86	5.66E-06	1.64	1.96E-02
Domain:card	2.91	8.14E-06	5.12	2.81E-02
Pirsf038545:chemokine receptor	1.42	7.28E-06	5.59	4.87E-03
Domain:leucine-zipper	1.65	1.92E-05	2.7	6.48E-02

Supplementary Table 3: Enrichment pathways for LPP gene

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Glycoprotein	13.18	1.68E-30	1.51	1.06E-27
Signal Peptide	13.18	1.14E-29	1.61	5.16E-26
Glycosylation Site:N-Linked (Glcnac...)	13.18	1.36E-28	1.51	6.16E-25
Extracellular matrix	13.18	4.46E-27	3.08	2.40E-24
Focal adhesion	13.18	5.36E-25	3.72	8.69E-23
Proteinaceous extracellular matrix	13.18	6.18E-25	3.07	3.33E-22
Cell adhesion	13.18	1.12E-24	2.32	3.87E-21
Biological adhesion	13.18	1.31E-24	2.32	4.53E-21
Extracellular region part	13.18	2.79E-21	1.98	1.50E-18
Enzyme linked receptor protein signaling pathway	13.18	1.16E-18	2.71	4.03E-15
Extracellular matrix part	13.18	2.47E-17	4.03	1.33E-14
Secreted	13.18	3.26E-15	1.63	2.04E-12
Extracellular structure organization	13.18	3.10E-14	3.26	1.08E-10
Ecm-receptor interaction	13.18	4.93E-14	4.27	7.99E-12
Short sequence motif:cell attachment site	13.18	1.48E-12	4.19	6.69E-09
Egf-like region, conserved site	13.18	2.37E-12	2.47	4.30E-09
Regulation of cell proliferation	13.18	1.25E-11	1.79	4.33E-08
Collagen	13.18	6.67E-09	5.16	3.59E-06
Adherens junction	9.85	4.56E-18	3.62	2.46E-15
Anchoring junction	9.85	4.13E-17	3.38	2.23E-14
Cell-substrate junction	9.85	2.85E-14	3.76	1.54E-11
Basolateral plasma membrane	9.85	4.90E-14	2.92	2.64E-11
Cell-substrate adhesion	9.85	1.67E-12	3.83	5.80E-09
Cell-substrate adherens junction	9.85	2.78E-12	3.6	1.50E-09
Cell-matrix adhesion	9.85	6.56E-11	3.75	2.27E-07
Cell junction assembly	9.85	1.38E-10	5.34	4.79E-07
Cell junction	9.85	4.43E-10	1.9	2.39E-07
Cell junction organization	9.85	6.67E-10	4.39	2.31E-06
Cell-cell junction	9.85	1.37E-06	2.22	7.40E-04

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Cell-substrate junction assembly	9.85	2.24E-05	4.99	7.46E-02
Proteinaceous extracellular matrix	5.96	6.18E-25	3.07	3.33E-22
Biological adhesion	5.96	1.31E-24	2.32	4.53E-21
Extracellular matrix organization	5.96	1.41E-19	4.61	4.89E-16
Extracellular structure organization	5.96	3.10E-14	3.26	1.08E-10
Ecm-receptor interaction	5.96	4.93E-14	4.27	7.99E-12
Short sequence motif:cell attachment site	5.96	1.48E-12	4.19	6.69E-09
Egf-like region, conserved site	5.96	2.37E-12	2.47	4.30E-09
Triple helix	5.96	7.45E-12	6.89	4.72E-09
Hydroxylation	5.96	1.34E-10	4.27	8.49E-08
Hydroxylysine	5.96	1.18E-09	6.16	7.49E-07
Cell binding	5.96	2.61E-09	8.99	1.65E-06
Von willebrand factor, type C	5.96	5.46E-09	4.99	9.88E-06
Trimer	5.96	6.38E-09	6.48	4.04E-06
Collagen	5.96	6.67E-09	5.16	3.59E-06
Collagen triple helix repeat	5.96	7.36E-09	3.5	1.33E-05
Hydroxyproline	5.96	3.02E-08	5.16	1.91E-05
Collagen	5.96	3.37E-08	3.31	2.13E-05
Domain:vwfc	5.96	5.01E-08	7.45	2.27E-04
Collagen fibril organization	5.96	9.98E-08	5.39	3.45E-04
Basement membrane	5.96	5.21E-07	4.61	3.30E-04
Von Willebrand factor, type A	5.96	2.75E-06	3.06	4.98E-03
Cell-substrate adherens junction	5.01	2.78E-12	3.6	1.50E-09
Domain:lim zinc-binding 3	5.01	4.35E-07	5.81	1.97E-03
Actin-binding	4.78	4.32E-11	2.59	2.73E-08
Cytoskeleton	4.78	1.47E-10	1.89	9.29E-08
Actin cytoskeleton organization	4.78	4.96E-10	2.49	1.72E-06
Actin filament-based process	4.78	1.99E-09	2.38	6.89E-06
Cell leading edge	4.78	1.28E-08	2.76	6.89E-06
Actin cytoskeleton	4.78	2.48E-08	2.16	1.34E-05
Cytoskeleton organization	4.78	3.58E-08	1.89	1.24E-04
Actin binding	4.78	2.53E-05	3.93	1.59E-02
Egf-like calcium-binding	4.73	1.24E-13	4.05	2.25E-10
Egf-like calcium-binding, conserved site	4.73	7.18E-13	3.94	1.30E-09
Domain:egf-like 2; calcium-binding	4.73	4.59E-08	4.12	2.08E-04
Cub	4.73	1.19E-05	3.5	2.13E-02
Regulation of cell morphogenesis	4.7	1.59E-06	2.55	5.48E-03
Regulation of cell shape	4.7	6.58E-06	3.48	2.25E-02
Gtpase activation	4.29	1.23E-08	2.76	7.80E-06
Regulation of small Gtpase mediated signal transduction	4.29	1.11E-06	2.07	3.83E-03
Egf-like calcium-binding	4.24	1.24E-13	4.05	2.25E-10
Egf-like calcium-binding, conserved site	4.24	7.18E-13	3.94	1.30E-09
Egf-like, type 3	4.24	8.00E-13	2.92	1.45E-09
Calcium	4.24	9.42E-13	1.87	5.96E-10

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Egf-type aspartate/asparagine hydroxylation conserved site	4.24	1.02E-12	3.9	1.85E-09
Egf-like domain	4.24	2.08E-12	2.78	1.32E-09
Egf-like region, conserved site	4.24	2.37E-12	2.47	4.30E-09
Egf calcium-binding	4.24	3.66E-12	4.38	6.62E-09
Egf-like	4.24	4.34E-11	2.72	7.86E-08
Domain:egf-like 4; calcium-binding	4.24	4.05E-10	6.09	1.83E-06
Domain:egf-like 5; calcium-binding	4.24	3.90E-09	5.43	1.77E-05
Domain:egf-like 3; calcium-binding	4.24	3.90E-09	5.43	1.77E-05
Domain:egf-like 1	4.24	4.48E-08	2.98	2.03E-04
Domain:egf-like 2; calcium-binding	4.24	4.59E-08	4.12	2.08E-04
Domain:egf-like 7; calcium-binding	4.24	1.36E-07	5.79	6.17E-04
Domain:egf-like 6; calcium-binding	4.24	4.35E-07	5.81	1.97E-03
Domain:egf-like 12; calcium-binding	4.24	1.09E-06	7.45	4.94E-03
Domain:egf-like 9; calcium-binding	4.24	2.30E-06	6.98	1.04E-02
Domain:egf-like 11; calcium-binding	4.24	7.41E-06	7.18	3.30E-02
Domain:egf-like 8; calcium-binding	4.24	1.23E-05	5.34	5.43E-02
Domain:egf-like 14; calcium-binding	4.24	2.40E-05	7.45	1.03E-01
Domain:egf-like 15; calcium-binding	4.24	2.40E-05	7.45	1.03E-01
Growth factor binding	4.24	2.59E-05	6.32	1.62E-02
Gtpase activation	4.15	1.23E-08	2.76	7.80E-06
Regulation of cell migration	4.1	1.89E-15	3.33	6.54E-12
Domain:egf-like 5; calcium-binding	4.04	3.90E-09	5.43	1.77E-05
Domain:egf-like 3; calcium-binding	4.04	3.90E-09	5.43	1.77E-05
Basement membrane	4.04	5.21E-07	4.61	3.30E-04
Basolateral plasma membrane	3.98	4.90E-14	2.92	2.64E-11
Ecm-receptor interaction	3.98	4.93E-14	4.27	7.99E-12
Cell-substrate adhesion	3.98	1.67E-12	3.83	5.80E-09
Cell-matrix adhesion	3.98	6.56E-11	3.75	2.27E-07
Integrin-mediated signaling pathway	3.98	4.50E-10	4.02	1.56E-06
Cell leading edge	3.98	1.28E-08	2.76	6.89E-06
Regulation of actin cytoskeleton	3.98	1.35E-07	2.24	2.19E-05
Hypertrophic cardiomyopathy (HCM)	3.98	9.31E-07	3.02	1.51E-04
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	3.98	2.40E-05	2.83	3.89E-03
Extracellular matrix part	3.94	2.47E-17	4.03	1.33E-14
Ecm-receptor interaction	3.94	4.93E-14	4.27	7.99E-12
Egf-like calcium-binding, conserved site	3.94	7.18E-13	3.94	1.30E-09
Egf-like region, conserved site	3.94	2.37E-12	2.47	4.30E-09
Basement membrane	3.94	1.10E-10	3.86	5.93E-08
Hydroxylysine	3.94	1.18E-09	6.16	7.49E-07
Cell binding	3.94	2.61E-09	8.99	1.65E-06
Domain:egf-like 3; calcium-binding	3.94	3.90E-09	5.43	1.77E-05
Trimer	3.94	6.38E-09	6.48	4.04E-06
Hydroxyproline	3.94	3.02E-08	5.16	1.91E-05
Domain:egf-like 1	3.94	4.48E-08	2.98	2.03E-04

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Basement membrane	3.94	5.21E-07	4.61	3.30E-04
Vasculature development	3.63	1.02E-20	3.2	3.54E-17
Blood vessel development	3.63	3.90E-20	3.19	1.35E-16
Regulation of cell motion	3.63	8.01E-16	3.19	2.69E-12
Skeletal system development	3.63	1.05E-15	2.62	3.46E-12
Regulation of cell migration	3.63	1.89E-15	3.33	6.54E-12
Tube development	3.63	3.21E-14	2.89	1.11E-10
Regulation of locomotion	3.63	1.59E-13	2.99	5.51E-10
Blood vessel morphogenesis	3.63	2.20E-13	2.87	7.62E-10
Branching morphogenesis of a tube	3.63	9.66E-12	4.49	3.34E-08
Regulation of cell proliferation	3.63	1.25E-11	1.79	4.33E-08
Angiogenesis	3.63	1.48E-11	3.1	5.13E-08
Positive regulation of cell motion	3.63	2.21E-10	3.51	7.66E-07
Morphogenesis of a branching structure	3.63	3.34E-10	3.95	1.16E-06
Positive regulation of cell migration	3.63	3.35E-10	3.63	1.16E-06
Respiratory tube development	3.63	7.01E-10	3.37	2.43E-06
Positive regulation of locomotion	3.63	1.03E-09	3.41	3.56E-06
Lung development	3.63	1.36E-09	3.37	4.71E-06
Respiratory system development	3.63	1.41E-08	3.09	4.87E-05
Tube morphogenesis	3.63	1.99E-08	2.87	6.89E-05
Urogenital system development	3.63	3.09E-07	2.84	1.07E-03
Embryonic morphogenesis	3.63	3.22E-07	2	1.11E-03
Palate development	3.63	4.60E-07	4.89	1.59E-03
Angiogenesis	3.63	6.41E-07	3.57	4.06E-04
Patterning of blood vessels	3.63	8.53E-07	5.96	2.95E-03
Regulation of mesenchymal cell proliferation	3.63	1.34E-06	6.37	4.64E-03
Heart development	3.63	2.32E-06	2.13	8.00E-03
Negative regulation of cell differentiation	3.63	2.63E-06	2.12	9.07E-03
Positive regulation of cell proliferation	3.63	6.18E-06	1.74	2.12E-02
Positive regulation of mesenchymal cell proliferation	3.63	7.89E-06	6.14	2.69E-02
Kidney development	3.63	8.61E-06	2.72	2.94E-02
Limb development	3.63	9.87E-06	2.63	3.36E-02
Appendage development	3.63	9.87E-06	2.63	3.36E-02
Pattern specification process	3.63	1.34E-05	1.91	4.54E-02
Appendage morphogenesis	3.63	1.51E-05	2.63	5.10E-02
Limb morphogenesis	3.63	1.51E-05	2.63	5.10E-02
Regulation of cell adhesion	3.57	4.78E-07	2.59	1.65E-03
Regulation of cell-substrate adhesion	3.57	5.13E-07	4.08	1.77E-03
Von Willebrand factor, type C	3.39	5.46E-09	4.99	9.88E-06
Trimer	3.39	6.38E-09	6.48	4.04E-06
Domain:vwfc	3.39	5.01E-08	7.45	2.27E-04
Domain:igfbp N-terminal	3.39	2.42E-05	5.59	1.04E-01
Positive regulation of cell motion	3.26	2.21E-10	3.51	7.66E-07
Positive regulation of cell migration	3.26	3.35E-10	3.63	1.16E-06
Positive regulation of locomotion	3.26	1.03E-09	3.41	3.56E-06

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Angiogenesis	3.26	6.41E-07	3.57	4.06E-04
Regulation of epithelial cell proliferation	3.26	1.74E-06	3.23	6.02E-03
Positive regulation of phosphorylation	3.26	1.04E-05	2.69	3.54E-02
Anchoring junction	3.23	4.13E-17	3.38	2.23E-14
Cell junction organization	3.23	6.67E-10	4.39	2.31E-06
Tube development	3.22	3.21E-14	2.89	1.11E-10
Branching morphogenesis of a tube	3.22	9.66E-12	4.49	3.34E-08
Morphogenesis of a branching structure	3.22	3.34E-10	3.95	1.16E-06
Transmembrane receptor protein serine/ threonine kinase signaling pathway	3.22	9.25E-10	3.34	3.20E-06
Tube morphogenesis	3.22	1.99E-08	2.87	6.89E-05
Tgf-beta signaling pathway	3.22	2.13E-08	3.3	3.45E-06
Chromosomal rearrangement	3.19	3.17E-05	1.89	1.98E-02
Domain:igfbp N-terminal	3.02	2.42E-05	5.59	1.04E-01
Growth factor binding	3.02	2.59E-05	6.32	1.62E-02
Domain:igfbp N-terminal	2.93	2.42E-05	5.59	1.04E-01
Repeat:spectrin 3	2.88	7.93E-07	6.09	3.58E-03
Calponin-like actin-binding	2.88	1.22E-06	3.3	2.21E-03
Domain:actin-binding	2.88	2.56E-06	6.14	1.15E-02
Repeat:spectrin 4	2.88	4.49E-06	5.85	2.01E-02
Repeat:spectrin 2	2.88	6.19E-06	5.16	2.77E-02
Repeat:spectrin 1	2.88	6.19E-06	5.16	2.77E-02
Domain:ch 2	2.88	9.61E-06	4.96	4.26E-02
Domain:ch 1	2.88	9.61E-06	4.96	4.26E-02
Actinin-type, actin-binding, conserved site	2.88	2.12E-05	5.02	3.76E-02
Domain:egf-like 2	2.8	1.92E-06	3.05	8.65E-03
Hypertrophic cardiomyopathy (HCM)	2.78	9.31E-07	3.02	1.51E-04
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	2.78	2.40E-05	2.83	3.89E-03

Supplementary Table 4: Enrichment pathways analysis in LPP-AS1

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Sexual reproduction	3.82	2.34E-15	3.39	4.35E-12
Male gamete generation	3.82	1.66E-14	3.96	3.08E-11
Spermatogenesis	3.82	1.66E-14	3.96	3.08E-11
Flagellum	2.96	3.49E-09	8.68	1.05E-06
Microtubule-based flagellum	1.38	1.88E-09	13.65	5.67E-07

Terms	Enrichment Score	P-value	Fold Enrichment	Bonferroni
Multicellular organism reproduction	3.82	1.98E-09	2.68	3.70E-06
Reproductive process in a multicellular organism	3.82	1.98E-09	2.68	3.70E-06
Reproductive cellular process	3.82	3.57E-07	3.76	6.66E-04
Ubl conjugation pathway	2.36	5.71E-06	2.05	2.04E-03
Spermatid development	3.82	2.77E-06	6.16	5.15E-03
Spermatid differentiation	3.82	4.84E-06	5.84	8.98E-03
Single fertilization	2.96	8.15E-06	5.54	1.51E-02
Modification-dependent macromolecule catabolic process	2.36	9.88E-06	2.08	1.82E-02
Modification-dependent protein catabolic process	2.36	9.88E-06	2.08	1.82E-02
Ubiquitin mediated proteolysis	2.36	1.99E-05	4.22	2.36E-03
Germ cell development	3.82	1.48E-05	4.12	2.72E-02
Fertilization	2.96	2.36E-05	4.56	4.30E-02
Proteolysis involved in cellular protein catabolic process	2.36	2.86E-05	1.99	5.19E-02